MASSACHUSETTS
INSTITUTE of TECHNOLOGY
LIBRARIES

DEWEY

HD28
.M414
no.3541-
93

WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

The Challenge: To Be Part of the Solution Instead of Being the Problem

February 1993 WP # 3541-93
CISL WP# 92-08

Stuart E. Madnick

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

The Challenge: To Be Part of the Solution Instead of Being the Problem

February 1993 WP # 3541-93
CISL WP# 92-08

Stuart E. Madnick

Stuart Madnick E53-321

Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 01239

M.I.T. LIBRARIES
MAR 2 5 1993

Reprinted from the *Proceedings of the Workshop on Information Technology & Systems* (WITS'92), December 1992, Dallas, Texas.

The Challenge: To Be Part of the Solution Instead of Being the Problem

Stuart E. Madnick
John Norris Maguire Professor of Information Technology and
Leaders for Manufacturing Professor of Management Science
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
(617) 253-6671
smadnick@eagle.mit.edu

Abstract: Society at large, and the Computer Science and Information Systems (CS/IS) fields in particular, are facing considerable challenges. Depending upon the directions that we take, CS/IS can either lead to solutions for these problems – or be a major contributor to the problems. The WITS community is uniquely positioned to address and solve problems that are outside of the traditional domains of either science and engineering or management. Examples are given of applications-driven technology-intensive research needed by business to support globalization, increased productivity, and rapid adaptation.

The Challenge: Symptoms

We can all take delight in the latest advances in information technologies: the newest high-speed chip, the increasingly exotic multi-media workstations, the incredibly sophisticated software, etc. Given all these great advances, it would seem that an enthusiastic future lies ahead. Instead, there are some very serious challenges facing the computer science/information systems (CS/IS) fields. Some of the symptoms are illustrated below.

Declining student enrollment. Nationwide percentage enrollment in CS/IS peaked around 1985-86 and has been dropping ever since. Some have claimed that it has dropped by as much as 75%. Although the specific numbers may be subject to debate, the drop seems clear.

Chief Executive Officer attitudes. Various reports have indicated that the majority of CEOs do not feel that they are getting sufficient value for the amount of information technology investment that they are making. Although there may be good reasons for outsourcing IT activities, for many CEOs it is primarily a way to get rid of something that it is unpleasant.

Student perceptions. There have been various studies made of career plans of students. In one study, three points were made regarding careers in CS/IS. First, the CS/IS fields were viewed as having good career opportunities. Second, CS/IS were viewed as fields in which one could make good

money. But, third, CS/IS were viewed as fields that were not fun! (That may explain why there are T.V. shows like "LA Law" but not "Chicago Information Technologist").

Economic analysis of the relationship between CS/IS and productivity. Several studies have indicated that, at the macroscopic level, there is no correlation between the increasing amount of investment in information technology and a corresponding increase in productivity (some studies have indicated a negative correlation.) Although we can argue about flaws in the simplicity of these studies, the fact remains that they do place the CS/IS fields on the defensive.

Taken separately or collectively, these factors do not indicate a healthy situation for attracting students or effectively impacting industry. Rather than leading to despair, these symptoms should trigger a reexamination of our existing assumptions and actions. The CS/IS fields have for years been major forces for reexamination and change in many fields and industries, such as financial services, manufacturing, transportation, etc. These "paradigm shifts" are almost always difficult and often resisted but do usually lead to new and more productive forms. It is now our time!

Although I will be presenting my own personal views on these issues, these concerns and the desire to seek effective responses has attracted the attention of others in recent years. One of the most recent and most comprehensive studies was reported in the book, *Computing the Future: A Broader Agenda for Computer Science and Engineering* by the National Research Council [HL92]. This paper makes use of some material from that study and the author's participation in related panel discussions.

The Challenge: Research Process

There are various ways in which research areas are identified and pursued. For example, in many engineering fields the following process is followed:

1. An application-specific problem is identified (e.g., aircraft wings develop cracks after a certain number of hours of flight).

2. The key aspects of the problem are abstracted and brought back to the laboratory for study (e.g., a vibration machine is used to simulate the affects of hours of flight).

3. Theories are developed and experiments are conducted to measure factors, explain reasons for material failures, and propose alternative materials or structures.

4. The results of this research are transferred back to industry to validate the conclusions and, once validated, to make use of these new solutions.

A disconnect from reality can take place if step 4 is omitted. Furthermore, over time a further disconnect can take place if step 1 is omitted and the input for the abstractions of step 2 comes primarily

(or only) from the already abstracted output of step 3. Eventually, the research being conducted and the results attained have less and less to do with the problems facing society at large. Thus, CS/IT efforts would not be providing benefits to industry and this ultimately would lead to a decrease in interest in CS/IT by students and executives – exactly the symptoms noted at the beginning of this paper. This concern also has been noticed by others. For example, the Association of Computing Machinery report on "The Scope and Directions of Computer Science: Computing, Applications, and Computational Science" [A91] warned:

> "A close interaction between computer researchers and others is essential so that the questions under investigation remain connected to real concerns. Otherwise computing research can drift into irrelevance and cease to earn public support. For this reason it is in the best interests of the computing profession for computer researchers to engage with applications."

Before going on, it should be noted that such abstracted research can be an important facet of our human search for knowledge and is a key element of what we often refer to as "basic science." The issue is not whether such approaches are "bad" but, rather, whether there are other approaches that can lead to more rapid solutions to important problems facing our society today.

Such a process is depicted in Figure 1. It starts by looking at the major problems facing businesses. Many of these problems can be resolved by applying existing solutions (this is often the role that consultants perform in helping companies to apply reported research results). But, certain problems may be new enough or difficult enough so that solutions do not currently exist. Furthermore, despite our enthusiasm, not all problems can be best solved through CS/IS research. From the list of hard problems, some important CS/IS research challenges can be found.

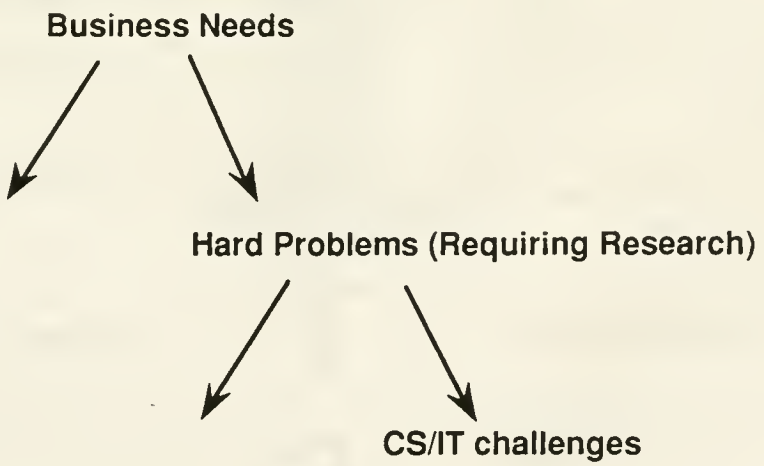

Figure 1. Process to Identify CS/IS Research Challenges

As one colleague noted, developing CS/IS-based information-intensive solutions to complex global transportation management problems, such as intransit visibility, can be just as challenging research as developing query optimization algorithms – but can have impacts that can be measured in billions of dollars of savings.

WITS Core

There is ambiguity and disagreements regarding the definitions and boundaries of fields such as computer science and information systems. Furthermore, given the rapid changes, the definitions are likely to continue to evolve over time. The CS/IS research agendas overlap with each other as well as with other fields in science and engineering as well as with various disciplines of management. On the other hand, some research issues have been largely neglected because management researchers felt that they were solely technical problems and thereby avoided them; at the same time, the science and engineering researchers felt that they were solely management problems and also avoided them; thus, nobody focused on them! In many ways, this segment, depicted in the center of Figure 2, represents the core of the WITS agenda. Research in this area requires a deep and enlightened understanding of both the management problems and the technologies.

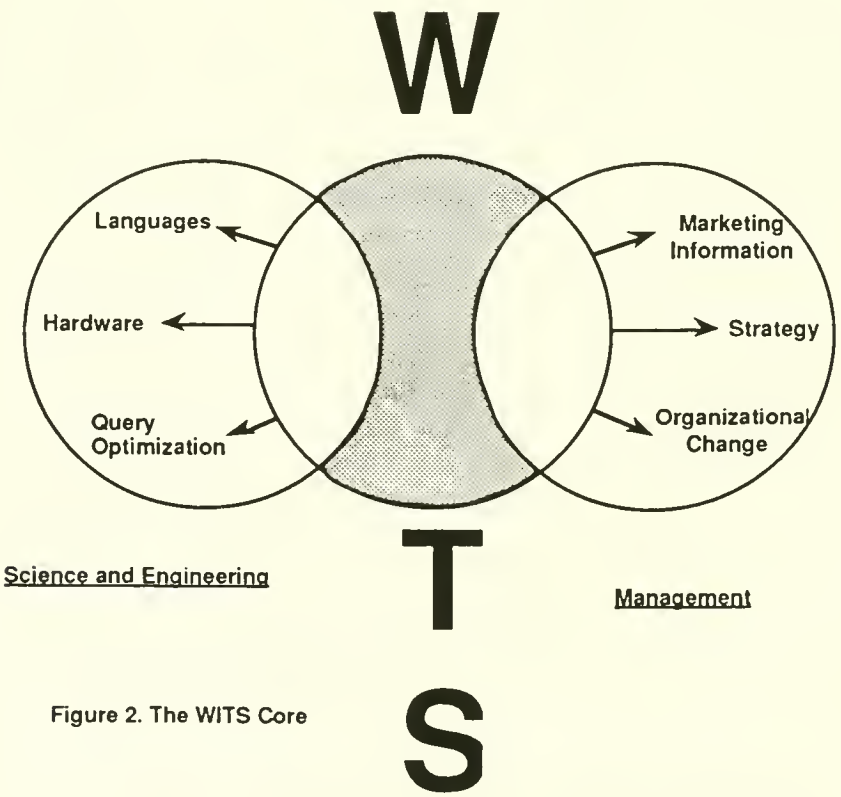

Figure 2. The WITS Core

Some researchers in the field of Information Systems (IS), such as Prof. Warren McFarlan of Harvard, have argued that although IS-related research may continue to expand, it will (or has) moved into the research agendas of the related functional areas, such as strategy (IT strategy), organizational studies (IT-based organizational change), and marketing (creation of enormous data bases), leaving the IS field empty. Without debating the correctness of this position that IS research will diffuse into existing management disciplines, the topics within the WITS core are unlikely to be "siezed" by any other other management discipline since they require significant technical research expertise.

In the remainder of this paper, we will identify some of the important business needs that are driving WITS core research.

The Challenge: Addressing Key Business Problems

Much of the content of this section emerged from this author's participation in the "Management in the 1990s" research project at MIT. This research project culminated with the publication of the book, *The Corporation of the 1990s: Information Technology and Organizational Transformation* (Edited by Michael Scott Morton) [S90].

Business today is characterized by dramatically increased geographic scope due to globalization and world-wide competition, severe time and productivity pressures, and a rapidly changing environment that often requires the restructuring of a company's operations in a very short time [M92]. For example, top management may need to restructure the organization from a company in which each site performs all of the functions relevant to the business, requiring only summary information from other offices, to a company in which each office is responsible for only specific functions, requiring much more information exchange from other offices. Movements toward consolidation (e.g., through bank mergers) are occuring at the same time as movements towards decentralization (e.g., the disintegration of IBM into multiple "Baby Blues").

Often the movement toward consolidation of previously independent companies or divisions into a single unit is motivated by the goal of gaining economy of scale. Thus, a rapid integration of separately maintained databases, processes, and networks would be necessary. Automated data conversion, interoperable network protocols, and transportable software systems are some of the major technological features necessary in such an environment.

Over and over the issue of time comes up. Popular phrases include "just-in-time", "continuous flow of information", "time-based functionality", and "time-to-market". Businesses are trying to compress the time from product concept to start of production, the time from product order to product ship, and the time to respond to a competitor's action. As an extreme case, many innovations in the financial services industry have a life span, from product concept to first deployment to final abandonment, of less than a month. In the business community, computer systems are often viewed as a major obstacle to time compression rather than a facilitator (one executive has a sign in his office that reads: "Once upon a time I thought that computers were the solution, now I realize that they are my major problem.")

One example that illustrates these needs and problems involves several ships full of goods being sent from the USA to help people in a foreign country suffering from a political disruption. While the ships were underway, the political situation improved in that country but a natural disaster struck another nearby country. The question posed was: Could one or more of the ships be diverted to help deal with the new disaster? This involved knowing what goods were supplied (they had come from multiple sources, each with their own computer systems), in which containers had those goods been stored, which containers were on which ships, and where was each of the ships currently located. Even though all of the necessary information existed in some computer system, there were so many disparate systems involved that a manual inspection of the contents of all the containers on each ship was ordered. Thus, a delay of several days occurred before the appropriate ships could be redeployed.

Although the above example may seem to be an interesting but isolated case, it represents much more the norm than the exception. For example, during the Gulf War, of the 40,000 containers of material shipped to the Gulf, about 28,000 had to be manually unloaded and inspected in order to determine their contents. In general, the physical movement of material was faster than the movement of the supporting information. Transportation and logistic systems, in general, represent major challenges to the effective utilization of information technology [Q91].

The Challenge: Key Research Issues

Effectively integrating information from multiple sources both within organizations and between organizations represent both an important solution to many critical business needs and a key challenge for CS/IS research [MSW90, MW91, SM89a]. An organization can be simultaneously "data rich" and "information poor" if they do not know how to identify, categorize, summarize, and organize the data. Three particular challenges are: data semantics acquisition, data quality, and data semantics evolution.

Data semantics acquisition. As business operations become increasingly dispersed geographically and functionally, differences in work processes at each site performed by people trained for each site will become more critical, leading to data incompatibilities and inconsistencies when these differing sites must interact. For example, a certain insurance company has 17 different definitions of the term "net written premium" which is their primary measure of sales. Which definition is used depends on which office and functional group within the company is using the term and for what purposes the definition is being used. A useful integration of this company's databases would need to reconcile conflicting definitions of terms when necessary. Before these differences could be reconciled, we would need to be able to represent the semantics of the data as used in each environment, this is sometimes called the "context" of the data [SM91b]. Further research on metadata definition and context representation would provide the basis for increasing the knowledge about data meanings and facilitate the automation of the data reconciliation and integration process [SM89b, SM91a].

Data quality. Organizations have become very concerned about quality in areas ranging from manufacturing quality to software program quality. Data quality, in comparison, has received relatively little attention. Issues relating to data quality are becoming increasingly important as information is moved through the organization. To a large extent, data quality considerations in the past were handled through personal familiarity; the user knew the characteristics of the data used in his or her organization and informally took this into account when using the data. This approach is not feasible as increasing numbers of information sources are used, many not well known to the user. They are increasingly exposed to data with various levels of quality for which they do not have first-hand familiarity. Furthermore, many current automated processes for converting, merging, and manipulating the data renders inaccessible information about the original data that might have conveyed information about its quality. For example, the source of a given piece of information is often a key element in judgements about its credibility and quality [WM90]. There are many additional data quality attributes that may be important, such as accuracy, completeness, timeliness, and stability. Defining and measuring the important data quality attributes and properly maintaining this quality-related information as data moves through and between systems represents a significant research challenge. With this quality information, decision makers will be better able to make effective use of the data.

Evolving semantics. It must be realized that autonomous databases are independently evolving in semantics as well as in context. For example, consider the situation of stock exchanges around the world. Not only are the stock prices changing continuously, but the definition of the stock price also can

change. At some time in the future, the Paris stock exchange will probably change from being measured in French francs to ECUs (European Currency Units). The normal "ticker tape" data feeds do not explicitly report the currency, it is implicit in the context of the source. More subtle examples include changes from reporting "last nominal price" to "last closing price" or from a percentage based pricing to actual prices, as currently happening at the Madrid stock exchange. Furthermore, in a historical database of stock prices, it must be recognized that the meanings had changed over time especially when doing a longitudinal analysis.

What is needed is not only a way to capture the meaning of each of these sources but also a way to represent the desired (or assumed) meaning of the receiver, which may be a human, an application, or another database. Then it would be possible to development context mediator algorithms to formally compare the semantics of the source and the receiver to determine if they are compatible, partially compatible, convertible, or incomparable. Research on the source/receiver model represents a direction towards solving this problem [SM91b].

Conclusions

It has been observed that the evolution of society goes through periodic discontinuities. At these transition points, new paradigms and fields of endeavor emerge. The WITS community is uniquely positioned to be a leader in helping to make this happen by effectively applying technology to critical problem areas. But, before we can help transform society, we have to be prepared to transform ourselves.

Acknowledgements

Work reported herein has been supported, in part, by the National Science Foundation, MIT's International Financial Services Research Center (IFSRC), MIT's Leaders For Manufacturing (LFM) Program, MIT's PROductivity From Information Technology (PROFIT) Program, and MIT's Total Data Quality Management (TDQM) Program.

REFERENCES

A91 Association for Computing Machinery, "The Scope and Directions of Computer Science: Computing, Applications, and Computational Science", *Communications of the ACM*, Volume 34(10 October 1991, p. 131.

HL92 Hartmanis, J. and Lin, H, (Editors), *Computing the Future: A Broader Agenda for Computer Science and Engineering*, National Academy Press, Washington, D.C., 1992.

M92 Madnick, S., "Chapter 16: Putting IT All Together Before it Falls Apart", in *Information Technology in Action: Trends and Perspectives*, Edited by Richard Y. Wang, Prentice-Hall, 1992.

MSW90 Madnick, S., Siegel, M. and Wang, Y.R., "The Composite Information Systems Laboratory (CISL) Project at MIT, *IEEE Data Engineering*, June 1990, pp. 10-15.

MW91 Madnick, S. and Wang, R., "Logical Connectivity: Applications, Requirements, and an Architecture", *Proceedings of the 1991 Hawaii International Conference on Systems Sciences.*

Q91 Quiddington, P., "Cruising Along the Information Highway", *MIT Management Magazine,* Fall 1991.

S90 Scott-Morton, M. S., (Editor), *The Corporation of the 1990s: Information Technology and Organizational Transformation,* Oxford University Press, 1990.

SM89a Madnick, S. et al, "CISL: Composing Answers from Disparate Information Systems, *Proceedings of the 1989 NSF Workshop on Heterogeneous Databases,* December 1989, Evanston, IL.

SM89b Siegel, M. and Madnick, S. , "Schema Integration Using Metadata", *Proceedings of the 1989 NSF Workshop on Heterogeneous Databases,* December 1989, Evanston, IL.

SM91a Siegel, M. and Madnick, S., "A Metadata Approach to Resolving Semantic Conflicts", *Proceedings of the VLDB Conference* (Barcelona, Spain), September 1991.

SM91b Siegel, M. and Madnick, S., "Context Interchange: Sharing the Meaning of Data", *SIGMOD Record,* December 1991, pp. 77-79.

WM90 Wang, R. and Madnick, S., "A Polygen Model for Heterogeneous Database Systems: The Source Tagging Perspective", *Proceedings of the VLDB Conference* (Brisbane, Australia), August 1990, pp. 519-538.

Date Due

Lib-26-67

MIT LIBRARIES DUPL

3 9080 00846561 6